

IN THE SPECIFICATION

On page 1, before line 3, insert the following.

--BACKGROUND OF THE INVENTION

Field of the Invention--

On page 1, before line 8, insert the following.

--Description of Related Art--

On page 2, before line 18, insert the following.

--SUMMARY OF THE INVENTION--

On page 3, before line 23, insert the following.

--BRIEF DESCRIPTION OF THE DRAWINGS--

On page 8, before line 19, insert the following.

--DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS--

Please amend the paragraph starting at page 21, line 17 and ending at page 22, line 21, as follows.

--If the count CNTBLW is greater than the predetermined number NHLD, then both the counts CNTABV and CNTBLW are reset in step S21 and the processing returns to step S5 where the control unit 86 waits, through the action of steps S3 and S5, for the next frame

which is above the detection threshold  $Th$ . If at step S19, the number of consecutive frames which are below the threshold is not greater than the predetermined number  $NHLD$ , then processing proceeds to step S23 where the frame number  $k$  is incremented. In step S25, the control unit 86 then determines if the bandpass modulation power  $w_k$  for the next frame is above the detection threshold  $Th$ . If it is not, then the processing returns to step S17, where the count  $CNTBL$   $CNTBLW$  of the number of consecutive frames below the threshold is incremented. If, on the other hand, the control unit 86 determines, in step S25, that the bandpass modulation power  $w_k$  for the next frame is above the detection threshold  $Th$ , then the processing passes from step S25 to step S27, where the number of frames which are below the detection threshold is reset to zero and the processing returns to step S7, where the number of frames which are above the detection threshold is incremented. Once the count  $CNTABV$  is above  $NDTCT$ , indicating speech has started, then the processing proceeds from step S9 to step S28, where the control unit 86 initiates the calculation of the start of speech point using a maximum likelihood calculation on recent frames. The state of the control unit 86 is then changed to be  $INSPEECH$  in step S29 and the processing returns to step S1.--

Please amend the paragraph starting at page 27, line 8 and ending at page 28, line 6, as follows.

--In this embodiment Laplacian statistics are used to model the noise and speech portions and the likelihood  $L_1$  that frames 1 to  $M$  in the buffer 92 are silence is given by:

$$L_2 = \left(2\sigma_1^2\right)^{-\frac{M}{2}} \exp\left(-\frac{\sqrt{2}}{\sigma_1} \sum_{i=1}^M |y_i|\right) \quad (2)$$

where  $y_i$  is the high-pass filtered energy and  $\sigma_1$  is the silence variance. Similarly, the likelihood  $L_2$  that frames  $M + 1$  to  $N$  are speech is given by:

---


$$L_1 = (2\sigma_2^2)^{-\frac{(N-M)}{2}} \exp\left(-\frac{\sqrt{2}}{\sigma_2} \sum_{i=M+1}^N |y_i - ay_{i-1}|\right) \quad (3)$$


---

$$L_2 = (2\sigma_2^2)^{-\frac{(N-M)}{2}} \exp\left(-\frac{\sqrt{2}}{\sigma_2} \sum_{i=M+1}^N |y_i - ay_{i-1}|\right) \quad (3)$$


---

where a first order auto-regressive process with a Laplacian driving term with variance  $\sigma_2$  has been used. The parameter  $a$  is the prediction co-efficient of the ~~auto-aggressive~~ auto-regressive model and, in this embodiment, a fixed value of 0.8 is used. The Laplacian statistics were found to be more representative of the data than the more usual Gaussian statistics and lead to more robust estimates and require less computation. However, Gaussian statistics can be used. Multiplying the likelihoods  $L_1$  and  $L_2$  gives the likelihood for a transition from silence to speech at frame  $M$ --

Please amend the paragraph starting at page 30, line 22 and ending at page 31, line 8, as follows.

--In the present embodiment, a mel spaced filter bank 69 having sixteen bands is used. The mel scale is well known in the art of speech analysis, and is a logarithmic scale that attempts to map the perceived frequency of a tone onto a linear scale. Figure 12 shows the output  $|\tilde{S}^k(f)|$   $|\tilde{S}^k(f')|$  of the mel spaced filter bank 69, when the samples shown in Figure 11 are

passed through the bank 69. The resulting envelope 100 of the magnitude spectrum is considerably smoother due to the averaging effect of the filter bank 69, although less so at the lower frequencies due to the logarithmic spacing of the filter bank.--

Please amend the paragraph starting at page 32, line 12 and ending at page 33, line 6, as follows.

--Figure 13 shows the envelope of the logged output from the mel filter bank 69, i.e.  $\log |\hat{S}^k(f)| |\hat{S}^k(f')|$ , which shows graphically the additive nature of two components 101 and 103. Component 101 is representative of the vocal tract characteristics, i.e.  $\log |V(f)|$ , and component 103 is representative of the excitation characteristics, i.e.  $\log |E(f)|$ . The peaks in component 101 occur at the formant frequencies of the vocal tract and the equally spaced peaks in component 103 occur at the harmonic frequencies of the pitch of the speaker.--

Please amend the paragraph starting at page 38, line 24 and ending at page 39, line 11, as follows.

--In addition to the nine cepstral coefficients mentioned above, the average energy of the speech signal within each frame is also used as a recognition feature for each input frame. Energy is an important feature since it can be used, among other things, to indicate whether or not the input speech signal during the frame corresponds to a voiced speech signal. As described above, the frame energy of each input frame is calculated in the energy calculation unit 76 and stored in buffer 78 shown in Figure 7. The energy for the current frame output by the buffer 78 is

then normalised by the normalising block 83 (Fig. 10) in order to remove the variation caused by variable recording conditions.--

On page 50, delete the text on line 1 and insert the following text on line 1.

--WHAT IS CLAIMED IS:--